What Is Claimed Is:

1.   A method of queuing InfiniBand receive traffic, comprising:

queuing one or more InfiniBand Send commands in a queue, wherein each said Send command comprises an encapsulated communication;

queuing a set of InfiniBand RDMA Read descriptors in said queue;

selecting an entry in said queue, wherein said entry comprises a Send command or a set of said RDMA Read descriptors;

if said selected entry is a set of said RDMA Read descriptors:

issuing a set of RDMA Read requests to retrieve portions of a communication described by said RDMA Read descriptors; and

as RDMA Read responses are received in response to said RDMA Read requests, assembling said described communication in said queue; and

forwarding a communication associated with said selected entry, for transmission on an external communication link, wherein said communication is one of:

said encapsulated communication if said selected entry is a Send command; and

said described communication if said selected entry is a set of RDMA Read descriptors.

2.   The method of claim 1, further comprising:

maintaining a single memory structure for queuing InfiniBand traffic received via multiple virtual lanes and multiple queue pairs, said single memory structure comprising said queue.

50

3.   The method of claim 2, wherein said queue comprises one or more linked lists of memory buffers within said single memory structure.

4.   The method of claim 1, further comprising:

maintaining an assembly area of said queue, in which said described communication is assembled; and

maintaining a queuing area of said queue, in which said one or more Send commands and said set of RDMA Read descriptors are queued.

5.   The method of claim 4, wherein said RDMA Read responses are placed directly into said assembly area upon receipt.

6.   The method of claim 4, further comprising, if said selected entry is a set of RDMA Read descriptors:

appending space to said assembly area of said queue based on an expected size of said described communication.

7.   The method of claim 4, further comprising:

maintaining a first head pointer configured to identify a head of said assembly area of said queue;

maintaining a first tail pointer configured to identify a tail of said assembly area of said queue;

maintaining a second head pointer configured to identify a head of said queuing area of said queue;

maintaining a second tail pointer configured to identify a tail of said queuing area of said queue; and

maintaining a next entry pointer configured to identify a next entry in said

51

queue to be processed after said forwarding.

8.   The method of claim 7, wherein said first head pointer is further configured to identify a beginning of said communication.

9.   The method of claim 7, wherein said second tail pointer is configured to identify where in said queue a next Send command or set of RDMA Read descriptors is to be queued.

10.   The method of claim 1, further comprising:
maintaining a set of pointers configured to identify a beginning and an end of said communication.

11.   The method of claim 10, wherein said set of pointers includes a head pointer configured to identify a head of said queue.

12.   The method of claim 1, wherein said assembling comprises dropping an RDMA Read response received out of order.

13.   A computer readable medium storing instructions that, when executed by a computer, cause the computer to perform a method of queuing InfiniBand receive traffic, the method comprising:
queuing one or more InfiniBand Send commands in a queue, wherein each said Send command comprises an encapsulated communication;
queuing a set of InfiniBand RDMA Read descriptors in said queue;
selecting an entry in said queue, wherein said entry comprises a Send command or a set of said RDMA Read descriptors;

52

if said selected entry is a set of said RDMA Read descriptors:

    issuing a set of RDMA Read requests to retrieve portions of a communication described by said RDMA Read descriptors; and

    as RDMA Read responses are received in response to said RDMA Read requests, assembling said described communication in said queue; and

forwarding a communication associated with said selected entry, for transmission on an external communication link, wherein said communication is one of:

    said encapsulated communication if said selected entry is a Send command; and

    said described communication if said selected entry is a set of RDMA Read descriptors.

14.   The computer readable medium of claim 13, wherein the method further comprises:

    maintaining an assembly area of said queue, in which said described communication is assembled; and

    maintaining a queuing area of said queue, in which said one or more Send commands and said set of RDMA Read descriptors are queued.

15.   The computer readable medium of claim 14, wherein the method further comprises, if said selected entry is one of said RDMA Read commands:

    appending space to said assembly area of said queue based on an expected size of said described communication.

16.   The computer readable medium of claim 15, wherein the method

further comprises:

maintaining a first head pointer configured to identify a head of said
assembly area of said queue;

maintaining a first tail pointer configured to identify a tail of said assembly
5   area of said queue;

maintaining a second head pointer configured to identify a head of said
queuing area of said queue;

maintaining a second tail pointer configured to identify a tail of said
queuing area of said queue; and

10   maintaining a next entry pointer configured to identify a next entry in said
queue to be processed after said forwarding.

17.   A method of queuing multiple types of traffic in a receive queue of
a communication interface, the method comprising:

15   queuing a first entry comprising a first communication forwarded to the
communication interface by a host;

queuing a second entry comprising a set of descriptors configured to
describe a second communication stored on the host;

processing said first entry, wherein processing said first entry comprises:

20   determining whether said first communication is complete; and

forwarding said first communication to a communication module
for transmission; and

processing said second entry, wherein processing said second entry
comprises:

25   issuing requests to obtain portions of said second communication
described by said descriptors;

assembling said second communication in said queue; and

54·

forwarding said second communication to the communication module for transmission.

18.   The method of claim 17, wherein processing said second entry further comprises:

determining whether said second communication has been fully assembled.

19.   The method of claim 17, further comprising:

maintaining a queuing area for queuing Send commands; and

maintaining an assembly area for assembling said second communication from said portions of said second communication.

20.   The method of claim 19, wherein processing said second entry further comprises:

placing said portions of said second communication directly into said assembly area upon receipt.

21.   The method of claim 17, further comprising:

forwarding a previous communication to the communication module; and

selecting whichever of said first entry and said second entry has been queued for the longest time.

22.   The method of claim 21, wherein said selecting comprises advancing a next entry pointer to the next entry in the receive queue.

23.   The method of claim 22, wherein said selecting further comprises

reading a portion of a payload of said next entry to determine a traffic type of said next entry.

24. The method of claim 17, wherein the receive queue comprises a set of linked memory buffers within a single memory structure configured as queues for one or more InfiniBand queue pairs.

25. The method of claim 24, wherein processing said second entry further comprises:

appending one or more free memory buffers of the single memory structure to the receive queue;

wherein said assembling comprises assembling said second communication in said one or more memory buffers.

26. A computer readable medium storing instructions that, when executed by a computer, cause the computer to perform a method of queuing multiple types of traffic in a receive queue of a communication interface, the method comprising:

queuing a first entry comprising a first communication forwarded to the communication interface by a host;

queuing a second entry comprising a set of descriptors configured to describe a second communication stored on the host;

processing said first entry, wherein processing said first entry comprises:

determining whether said first communication is complete; and

forwarding said first communication to a communication module for transmission; and

processing said second entry, wherein processing said second entry

56

comprises:

issuing requests to obtain portions of said second communication described by said descriptors;

assembling said second communication in said queue; and

5   forwarding said second communication to the communication module for transmission.

27.   The computer readable medium of claim 26, wherein the method further comprises:

10   maintaining a queuing area for queuing Send commands; and

maintaining an assembly area for assembling said second communication from said portions of said second communication.

28.   The computer readable medium of claim 27, wherein processing

15 said second entry further comprises:

placing said portions of said second communication directly into said assembly area upon receipt.

29.   An apparatus for queuing multiple types of receive traffic in a

20 communication interface, comprising:

a queue for queuing multiple types of receive traffic commands, wherein each said command is associated with a communication to be transmitted from the communication interface;

a head pointer configured to identify a head of said queue;

25   a tail pointer configured to identify a tail of said queue, wherein said traffic commands are enqueued at said tail; and

a next entry pointer configured to identify a next entry in said queue to be

processed.

30.   The apparatus of claim 29, wherein said queue comprises an
assembly area for assembling a communication associated with a first type of
5   receive traffic command.

31.   The apparatus of claim 30, wherein said queue further comprises a
queuing area for queuing a second type of receive traffic command.

10   32.   The apparatus of claim 31, wherein said assembly area and said
queuing area are each delimited by a head pointer and a tail pointer.

33.   The apparatus of claim 30, wherein said first type of receive traffic
command is an InfiniBand Send command comprising a set of RDMA read
15   descriptors configured to identify the communication associated with said first
type of receive traffic command.

34.   The apparatus of claim 33, wherein a second type of receive traffic
command is an InfiniBand Send command configured to encapsulate the
20   communication associated with said second type of receive traffic command.

35.   The apparatus of claim 30, wherein:
said first type of receive traffic command comprises a set of descriptors,
wherein each said descriptor is configured to describe a portion of the
25   communication associated with said command; and
the apparatus is configured to issue read requests to retrieve the portions of
the communication described by the set of descriptors and assemble said portions

58

in said assembly area.

36.   The apparatus of claim 29, further comprising:

a transmit module configured to transmit the communications associated

5   with said receive traffic commands;

wherein each communication associated with a receive traffic command is

forwarded from said queue to said transmit module after the communication is

determined to be complete.

10   37.   The apparatus of claim 36, wherein a communication is forwarded

from said queue to said transmit module by passing to the transmit module a set of

pointers delimiting the communication within said queue rather than passing the

communication.

15   38.   The apparatus of claim 29, wherein said queue comprises one or

more linked lists of buffers within a memory structure configured to queue receive

traffic for multiple communication connections.

39.   A method of maintaining ordering of transmission of outbound

20   communications from an InfiniBand channel adapter, the method comprising:

receiving on a first queue pair a first InfiniBand packet payload comprising

a set of RDMA (Remote Direct Memory Access) Read descriptors describing a

first communication;

after receiving said first InfiniBand packet, receiving on the first queue

25   pair a second InfiniBand packet payload comprising a portion of a second

communication;

after receiving said second InfiniBand packet, processing said first

InfiniBand packet payload by:

dispatching RDMA Read requests corresponding to said set of RDMA Read descriptors;

receiving responses to said RDMA Read requests, said responses comprising portions of the first communication;

assembling the first communication; and

transmitting the first communication from the channel adapter; and

only after said processing said first InfiniBand packet, processing said second InfiniBand packet.

40. The method of claim 39, wherein said processing said second InfiniBand packet comprises transmitting the second communication from the channel adapter.

41. The method of claim 39, wherein:

said receiving a first InfiniBand packet payload comprises queuing said first InfiniBand packet payload in a first portion of a queue associated with the first queue pair; and

said receiving a second InfiniBand packet payload comprises queuing said second InfiniBand packet payload in the first portion of the queue.

42. The method of claim 41, wherein said assembling comprises:

assembling said portions of the first communication in a second portion of the queue.

                    Inventors: Mott, *et al.*